Adaptive Optimal Observer Design via Approximate Dynamic Programming

Jing Na, *Member IEEE,* Guido Herrmann, *Senior Member IEEE*, Kyriakos G. Vamvoudakis, *Senior Member IEEE*

***Abstract*—This paper presents an optimal observer design framework using a recently emerging method, approximate dynamic programming (ADP), to minimize a predefined cost function. We first exploit the duality between the linear optimal observer and the linear quadratic tracking (LQT) control. We show that the optimal observer design can be formulated as an optimal control problem subject to a specific cost function, and thus the solution can be obtained by solving an algebraic Riccati equation (ARE). For nonlinear systems, we further introduce an optimal observer design formulation and suggest a modified policy iteration method. Finally, to solve the problem online we propose a framework based on ADP and specifically on an approximator structure. Namely, a critic approximator is used to estimate the optimal value function, and a newly developed tuning law is proposed to find the parameters online. The stability and the performance are guaranteed with rigorous proofs. Numerical simulations are given to validate the theoretical studies.**

***Index Terms*—Optimal observer design, approximate dynamic programming (ADP), policy iteration.**

I. Introduction

In advanced control system design, a critical assumption is that the system states are available. This requirement, however, may not be always true in practice due to limited transducer costs, partial observability or even sensor noise. This fact has triggered extensive research on observer design, which can reconstruct the system states by means of limited output measurements. In the seminal work [1], Luenberger proposed a closed-loop observer for linear systems, where the output error is added to the observer to guarantee stability. This idea was also extended to nonlinear systems [2]. In the past decades, other observer design methodologies, e.g. high-gain observers [3], robust observers [4], sliding mode observers [5, 6] and adaptive observers [7], have also been suggested.

The observer usually has the same dynamics as the actual system, while appropriate compensators/corrections (e.g. output error or sliding mode term) are added to drive the observer states to the actual states. However, most observers are not optimal in the sense of minimizing a predefined performance index [8]. In this respect, the notable Kalman filter [9] was proposed to minimize the error covariance by choosing an appropriate feedback gain. Specifically, the duality of the Kalman filter and the linear quadratic regulation (LQR) was studied in [9]. Another interesting alternative derivation of the Kalman filter can be found in [10] in terms of optimal control theory [11]. An optimal filter was also studied [12] by considering a State Dependent Riccati Equation (SDRE). Recent work in [8, 13] investigated the duality of the optimal observer design and the linear quadratic tracking (LQT) control, and suggested a novel observer design methodology to minimize a quadratic performance functional with respect to the observer output error and the correction action. In this optimal observer synthesis, an algebraic Riccati equation (ARE) should be solved offline, which creates difficulties for the online implementation. In fact, although optimal control theory has been well developed in the past decades, optimal observer design has been rarely studied beyond Kalman filter.

The aim of this paper is to study an optimal observer design based on approximate dynamic programming (ADP) [14]. We first formulate the observer design as an optimal output tracking control problem following their duality property [8, 13]. Then, for linear systems, we derive an optimal observer solution by means of the principle of optimality [11]. This eventually relies on the solution of a standard ARE. The idea is then further extended to nonlinear systems. We introduce a specific optimal observer formulation for nonlinear systems based on ADP [15, 16]. Then, an offline policy iteration method is proposed to solve the nonlinear optimization equations by further extending the ideas of Kleinman's method [20] and modifying the policy iteration approach [21, 22]. Finally, we use a critic function approximator to estimate the optimal cost value to online solve the derived observer Hamilton-Jacobi-Bellman (HJB) equation [17]. Finally a new tuning law [18] is derived to estimate the critic parameters, along with an appropriate stability proof. Simulations are given to show the validity of the suggested algorithms.

II. Linear Optimal Observer Design Based on Linear Quadratic Tracking Control Method

Consider the linear system of the form

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases} \tag{1}$$

where $x \in \mathbb{R}^n$ is the unknown system state, $y \in \mathbb{R}^p$ is the measured output, and $u \in \mathbb{R}^m$ is the system input, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$ are system matrices. It is assumed that the pair (A, C) is observable and (A, B) is controllable.

The problem to be studied is to design an optimal observer for system (1), such that the state x can be reconstructed, while a performance index is minimized.

Remark 1: It is interesting to find that in [8, 13] the optimal observer design for a particular kind of linear systems (i.e. the system input is assumed to be zero, $u = 0$) can be considered

*This work was supported by the Marie Curie Intra-European Fellowships Project AECE under Grant FP7-PEOPLE-2013-IEF-625531, National Natural Science Foundation of China (NSFC) (No. 61573174), and Newton Mobility Grant jointly funded by NSFC (No. IE150833/ 6151101245).

Jing Na is with the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming, China. (*Corresponding author*. e-mail: najing25@163.com).

Guido Herrmann is with the Department of Mechanical Engineering, University of Bristol, BS8 1TR, UK (e-mail: g.herrmann@bristol.ac.uk).

Kyriakos G. Vamvoudakis is with the Kevin T. Crofton Department of Aerospace and Ocean Engineering, Virginia Tech (email: kyriakos@vt.edu).

as a standard linear quadratic tracking (LQT) control problem [17], [19], and thus solved by using optimal control theory.

The aim of this section is to show that the idea of LQT control design can be extended to design an optimal observer by considering the duality between the LQT and linear observer design. This can be achieved by further modifying the observer design in [8, 13], where nonzero control input is considered. In this case, the observer can be considered as a system which tries to track the output of the original system.

Thus, the observer should have the same dynamic equation as that of the plant. Hence, we design an observer as

$$\begin{cases} \dot{\hat{x}} = A\hat{x} + Bu + B\bar{u} \\ \hat{y} = C\hat{x} \end{cases} \tag{2}$$

where $\hat{x}$ is the estimate of x, $\hat{y}$ is the observer output, and $\bar{u} \in \mathbb{R}^m$ is the correction term.

Remark 2: It is known that the observer response is improved using the information of the system output, thus the correction term $\bar{u}$ in (2) is analogous to a control for a system.

Thus, we can now design an 'optimal control', $\bar{u}$, which minimizes the following cost function

$$V(y,\hat{x}) = \frac{1}{2}\int_t^\infty [(\hat{y}-y)^T Q(\hat{y}-y) + \bar{u}^T R\bar{u}]d\tau, \forall x \text{ and } t > 0 \tag{3}$$

where $Q \in \mathbb{R}^{p\times p}$ and $R \in \mathbb{R}^{m\times m}$ are positive definite matrices chosen to tradeoff the performance and correction effort as justified in optimal control [17].

Hence, the problem can be formulated as: to find a 'control' $\bar{u}$ to minimize the cost function (3) subject to constraint (2) for $\forall x, u$. Clearly, we can solve this problem in terms of optimal control theory [17].

We can now define the Hamiltonian as

$$H(\hat{x},\bar{u},\lambda) = \lambda^T[A\hat{x} + Bu + B\bar{u}] + \frac{1}{2}[(\hat{y}-y)^T Q(\hat{y}-y) + \bar{u}^T R\bar{u}] \tag{4}$$

where $\lambda \in \mathbb{R}^n$ denotes the adjoint variable.

Then the necessary conditions for optimality are given by

$$\frac{\partial H}{\partial \bar{u}} = 0, \tag{5}$$

$$\dot{\lambda} = -\frac{\partial H}{\partial \hat{x}}. \tag{6}$$

By solving (5) and (6) along (4), one has

$$\bar{u} = -R^{-1}B^T\lambda, \tag{7}$$

$$\dot{\lambda} = C^T Q(y-\hat{y}) - A^T\lambda. \tag{8}$$

To obtain the adjoint variable, the 'sweep' method [11] can be used by setting

$$\lambda = P\hat{x} - g, \tag{9}$$

where $P \in \mathbb{R}^{n\times n}$, $g \in \mathbb{R}^n$ are the influence matrix and function. It is shown that the correction term (7) with (9) consists of a linear feedback term plus an additional feedforward term. Then, from (2), (7) and (9) we have that

$$\begin{aligned} \dot{\lambda} &= \dot{P}\hat{x} + P\dot{\hat{x}} - \dot{g} = \dot{P}\hat{x} + P(A\hat{x} + Bu + B\bar{u}) - \dot{g} \\ &= \dot{P}\hat{x} + P[A\hat{x} + Bu - BR^{-1}B^T(P\hat{x} - g)] - \dot{g} \end{aligned}. \tag{10}$$

On the other hand, by combining (8) and (9) we can find

$$\dot{\lambda} = C^T Q(y - C\hat{x}) - A^T(P\hat{x} - g). \tag{11}$$

Hence, a feasible solution of above equations (10) and (11) can be given as

$$\dot{P} + PA + A^T P + C^T QC - PBR^{-1}B^T P = 0, \tag{12}$$

$$\dot{g} + (A^T - PBR^{-1}B^T)g + C^T Qy - PBu = 0. \tag{13}$$

Since we consider the optimal observer design over infinite time horizon, the observer gains tend to constants as shown in [11], i.e. $\dot{P}, \dot{g} \to 0$. Consequently, the feedback gain P in (12) can be obtained by solving a standard ARE

$$PA + A^T P + C^T QC - PBR^{-1}B^T P = 0 \tag{14}$$

and the feedforward term g in (13) is given by

$$g = -(A^T - PBR^{-1}B^T)^{-1}(C^T Qy - PBu). \tag{15}$$

Remark 3: In the above analysis, we have shown the duality of the optimal control and the optimal observer, that is the optimal observer design in (2) to find an optimal compensator $\bar{u}$ and to minimize the cost function (3) is indeed equivalent to the LQT control design for (2) with the cost function (3). Thus, the optimal observer design is solved by extending the standard solution for the optimal LQT control problem [17].

To implement the observer (2), we have to solve a standard ARE (14) for P and calculate (15) for g, respectively. However, solving the ARE numerically may lead to computational complexity although some commercial software (e.g. Matlab) can be used in an offline manner. Moreover, the above analysis is for linear systems only.

III. NONLINEAR OPTIMAL OBSERVER BASED ON ADP

In this section, we will generalize the idea of LQT to design an optimal observer for nonlinear systems. Moreover, we will present an alternative method to solve the optimal equations by using the idea of approximate dynamic programming.

Consider the following nonlinear system

$$\begin{cases} \dot{x} = Ax + Bf(x,u) \\ y = Cx \end{cases} \tag{16}$$

where $x \in \mathbb{R}^n$ is the unknown system state, $y \in \mathbb{R}^p$ is the measured output, and $u \in \mathbb{R}^m$ is the system input, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$ are system matrices, and $f(x,u) \in \mathbb{R}^m$ is a nonlinear function. It is assumed that $f(x,u)$ is Lipschitz continuous and $\partial f(x,u)/\partial x|_{x=0} = 0$, (A,C) is observable and (A,B) is controllable. Thus, system (16) is observable.

Similarly, we design an observer for system (16) as

$$\begin{cases} \dot{\hat{x}} = A\hat{x} + Bf(\hat{x},u) + B\bar{u} \\ \hat{y} = C\hat{x} \end{cases} \tag{17}$$

where $\hat{x}$ is the observer state, $\hat{y}$ is the observer output, and $\bar{u} \in \mathbb{R}^m$ is the correction term to address the nonlinearities and to minimize the following infinite-horizon cost function

$$V = \frac{1}{2}\int_t^\infty [(\hat{y}-y)^T Q(\hat{y}-y) + \bar{u}^T R\bar{u}]d\tau, \forall x \text{ and } t > 0 \tag{18}$$

where $Q \in \mathbb{R}^{p\times p}$ and $R \in \mathbb{R}^{m\times m}$ are positive definite matrices.

A. Nonlinear Optimal Observer Design

Similar to the analysis for linear systems, the correction

term $\bar{u}$ in (17) is analogous to the control input of a system. From this perspective, the observer design in (17) can be taken as an optimal control problem for (17), which is formulated such that we eventually find a 'control' $\bar{u}$ to minimize the cost function (18) subject to the constraint given by (17).

In this respect, we can solve this optimization problem again in terms of optimal control theory [17]. We differentiate the value function (18) and write the Bellman equation as

$$\dot{V} = -\frac{1}{2}[(\hat{y}-y)^T Q(\hat{y}-y) + \bar{u}^T R\bar{u}]. \quad (19)$$

Then we can define the Hamiltonian as

$$H(\hat{x},\bar{u},V) = V_x^T(A\hat{x} + B\hat{f} + B\bar{u}) + \frac{1}{2}[(\hat{y}-y)^T Q(\hat{y}-y) + \bar{u}^T R\bar{u}], \forall x,u \quad (20)$$

where $V_x \triangleq \partial V / \partial \hat{x}$ denotes the partial derivative of the cost function $V(\hat{x})$ with respect to $\hat{x}$, and $\hat{f} = f(\hat{x},u)$.

We are interested to find the optimal cost function $V^*(\hat{x})$ with any $\bar{u}$ within admissible control set $\Psi(\Omega)$ [17]

$$V^* = \min_{\bar{u}\in\Psi(\Omega)}\left(\frac{1}{2}\int_t^\infty [(\hat{y}-y)^T Q(\hat{y}-y) + \bar{u}^T R\bar{u}]d\tau\right), \forall x \quad (21)$$

which satisfies the HJB equation

$$0 = \min\left[H(\hat{x},\bar{u}^*,V^*)\right] = V_x^{*T}[A\hat{x} + B\hat{f} + B\bar{u}^*] + \frac{1}{2}[(\hat{y}-y)^T Q(\hat{y}-y) + \bar{u}^{*T} R\bar{u}^*], \quad \forall x,u \quad (22)$$

Then from the stationarity condition, the ideal optimal action $\bar{u}^*$ is derived by solving $\partial H(\hat{x},y,\bar{u}^*,V^*)/\partial\bar{u}^* = 0$ as

$$\bar{u}^* = -R^{-1}B^T\frac{\partial V^*}{\partial \hat{x}}. \quad (23)$$

We are now ready to present the following result:

Theorem 1: Consider the observer (17) subject to the cost function (18). Then the optimal compensator (23) guarantees that the observer (17) is stable, and $\hat{x}$ converges to x for $t \to \infty$.

Proof: The derivative of the value function $V(\hat{x})$ with respect to t and (17) is given as

$$\frac{dV}{dt} = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial \hat{x}}\dot{\hat{x}} = V_x^T(A\hat{x} + B\hat{f} + B\bar{u}). \quad (24)$$

From (20) and (24) we have

$$H(\hat{x},y,\bar{u},V) = \frac{dV}{dt} + \frac{1}{2}[(\hat{y}-y)^T Q(\hat{y}-y) + \bar{u}^T R\bar{u}]. \quad (25)$$

For the optimal solution $\bar{u} = \bar{u}^*$, the value function V^* satisfies the HJB equation, i.e. $H(\hat{x},y,\bar{u}^*,V^*) = 0$, thus we know that

$$\frac{dV^*}{dt} = -\frac{1}{2}[(\hat{y}-y)^T Q(\hat{y}-y) + \bar{u}^{*T} R\bar{u}^*]. \quad (26)$$

Integrating both sides of above equation, it yields

$$V^*(t) - V^*(0) = -\frac{1}{2}\int_0^t [(\hat{y}-y)^T Q(\hat{y}-y) + \bar{u}^{*T} R\bar{u}^*]d\tau \le 0. \quad (27)$$

In this case, the optimal cost value V^* and thus $\hat{y}-y$ are bounded. Moreover, one may further use LaSalle's extension and show that V^* is asymptotically convergent. Then based on (23), we know that $\bar{u}^*$ will converge to zero. Thus, we consider that the pair (A,C) is observable and $f(x,u)$ is Lipschitz continuous, and have that $\hat{y}-y$ asymptotically converges to zero. Then according to Proposition 2.1 of [24], the observer states $\hat{x}$ will converge to x asymptotically. ◇

B. Offline Solution via Policy Iteration

Theoretically, the optimal observer can be synthesized from (23). However, the solution may not be obtained directly using (22) and (23) because the optimal value function V^* is derived by solving the HJB equation (22). This is extremely difficult by means of analytical approaches. Inspired by Kleinman's algorithm [20], we first present an offline policy iteration method to obtain the approximated solution of HJB equation (22).

Algorithm 1-Offline Policy Iteration for HJB Equation

1: **Start procedure**

2: **Initialization**: Start with a correction term $\bar{u}^0$, which stabilizes the observer (17), and set i=1

3: **while** $\left\|V^{i+1} - V^i\right\| \ge \varepsilon$ for a small threshold $\varepsilon > 0$, **do**

i) Policy evaluation: Find the cost function V^i using the Bellman equation

$$V_x^{iT}(A\hat{x} + B\hat{f} + B\bar{u}^i) + \frac{1}{2}[(\hat{y}-y)^T Q(\hat{y}-y) + \bar{u}^{iT} R\bar{u}^i] = 0 \quad (28)$$

4: **ii) Policy improvement:** Update the policy by

$$\bar{u}^{i+1} = -R^{-1}B^T V_x^i \quad (29)$$

5: **iii) Iteration:** let $i := i+1$

6: **end while**

7: **end procedure**

The above Algorithm 1 extends the offline policy iteration algorithm in [21] for observer designs of nonlinear systems. The following lemma proved in [21] shows that the policies are stable if the initial policy is admissible.

Lemma 1 [21]: If the initial correct $\bar{u}^0$ in Algorithm 1 is stable, we know: 1) $\bar{u}^i$ is stable; 2) $V^* \le V^{i+1} \le V^i$ with V^* being the optimal solution of HJB equation (22); 3) $\lim_{i\to\infty} V^i = V^*$, and $\lim_{i\to\infty}\bar{u}^i = \bar{u}^*$. ◇

C. Online Solution via ADP

We will use the idea of ADP and present an online solution for the observer design. The basic idea is to introduce an online approximator for the optimal value function given in (21), which is continuous on a compact set Ω. Thus, a critic function approximator for $V^*(X)$ is defined as [21, 22]

$$V^*(X) = W^T\phi(X) + \varepsilon_v, \forall X = [\hat{x}^T, y^T]^T \in \mathbb{R}^{n+p} \quad (30)$$

and its derivative with respect to $\hat{x}$ is

$$\frac{\partial V^*(X)}{\partial \hat{x}} = \nabla\phi^T W + \nabla\varepsilon_v \quad (31)$$

where $W=[W_1,\cdots,W_l]^T\in\mathbb{R}^l$ is the unknown parameter vector, $\phi(X)=[\phi_1,\cdots,\phi_l]^T\in\mathbb{R}^l$ is the regressor with $l>0$, and ε_v denotes the approximation error. Moreover, $\nabla\phi=\partial\phi/\partial\hat{x}$ and $\nabla\varepsilon_v=\partial\varepsilon_v/\partial\hat{x}$ define the partial derivative of ϕ and ε with respect to $\hat{x}$.

Assumption 1 [15, 16, 21]: The critic parameter W and the regressor functions ϕ, $\nabla\phi$ are bounded by $\|W\|\le W_N$, $\|\phi\|\le\phi_N$, $\|\nabla\phi\|\le\phi_M$; the approximation errors ε_v, $\nabla\varepsilon_v$ are also bounded as $\|\nabla\varepsilon_v\|\le\phi_\varepsilon$. ◇

Since the ideal parameter W is unknown, a practical critic approximator $\hat{V}(X)$ for estimating $V^*(X)$ is used

$$\hat{V}(X)=\hat{W}^T\phi(X) \tag{32}$$

where $\hat{W}$ denotes the estimated parameter of W.

Using (32), we can rewrite $\bar{u}$ as

$$\bar{u}=-R^{-1}B^T\frac{\partial\hat{V}}{\partial\hat{x}}=-R^{-1}B^T\nabla\phi^T(X)\hat{W}. \tag{33}$$

The regressor $\{\phi_i(X):i=1,\cdots,l\}$ can be selected so that all elements of $\phi(X)$ are linearly independent [21, 22]. Thus, based on the Weierstrass approximation theorem, we know that V^* and $\partial V^*/\partial\hat{x}$ can be represented as (30)-(31) with ε_v, $\nabla\varepsilon_v\to 0$ for $l\to+\infty$, and thus can be estimated by (32).

The remaining problem is to find an adaptive algorithm to obtain online the estimated parameter $\hat{W}$, which converges to W. For any fixed policy $\bar{u}$, the approximated Bellman equation with (31) can be given as

$$H=W^T\nabla\phi\left(A\hat{x}+B\hat{f}+B\bar{u}\right)+\frac{1}{2}[(\hat{y}-y)^TQ(\hat{y}-y)+\bar{u}^TR\bar{u}]=\varepsilon_{HJB} \tag{34}$$

where $\varepsilon_{HJB}=-\nabla\varepsilon_v(A\hat{x}+B\hat{f}+B\bar{u})$ is the residual error due to the critic approximation errors ε_v, $\nabla\varepsilon_v$. We know $\varepsilon_v,\nabla\varepsilon_v\to 0$ for $l\to+\infty$ and thus ε_{HJB} is bounded [21].

To design an adaptive law for estimating W, we denote $\Xi=\nabla\phi\left(A\hat{x}+B\hat{f}+B\bar{u}\right)$ and $\Theta=\frac{1}{2}[(\hat{y}-y)^TQ(\hat{y}-y)+\bar{u}^TR\bar{u}]$. Thus, the Bellman equation (34) can be written as

$$\Theta=-W^T\Xi+\varepsilon_{HJB}. \tag{35}$$

One can find from (35) that the unknown parameter W is now linearly parameterized. Therefore, the idea originally presented in our previous work [18] for designing adaptive laws can be further extended. We denote the auxiliary matrix $P_2\in\mathbb{R}^{l\times l}$ and vector $Q_2\in\mathbb{R}^l$ as

$$\begin{cases}\dot{P}_2=-\ell P_2+\Xi\Xi^T, & P_2(0)=0\\ \dot{Q}_2=-\ell Q_2+\Xi\Theta, & Q_2(0)=0\end{cases} \tag{36}$$

with $\ell>0$ being a constant forgetting factor. Note the above equation can be easily implemented to obtain P_2 and Q_2 by using a low pass filter.

Then a new auxiliary vector $M\in\mathbb{R}^l$ can be obtained by

$$M=P_2\hat{W}+Q_2. \tag{37}$$

where $\hat{W}$ is the estimation of the parameter W, which can be updated by

$$\dot{\hat{W}}=-\Gamma M \tag{38}$$

where $\Gamma>0$ is a constant learning gain.

The following definition is needed before we state the main results of this section.

Definition 1: A function vector ϕ is persistently excited (PE) if there exist constants $\tau>0,\varepsilon>0$ such that $\int_t^{t+\tau}\phi(r)\phi^T(r)dr\ge\varepsilon I,\ \forall t\ge 0$.

Now, we have the following results:

Theorem 2: Consider the critic approximator in (32) and the adaptive law in (38). Assume that the vector Ξ in (35) is PE, then one has:

i) when the approximation errors are zero (i.e. $\varepsilon_v,\nabla\varepsilon_v=0$ and thus $\varepsilon_{HJB}=0$), $\tilde{W}$ converges to zero exponentially, and the approximated policy $\bar{u}$ in (33) will converge to the optimal solution $\bar{u}^*$ given by (23).

ii) when $\varepsilon_v,\nabla\varepsilon_v\ne 0$ and thus $\varepsilon_{HJB}\ne 0$, then $\tilde{W}$ converges to a small compact set around zero, and $\bar{u}$ converges to a neighborhood around $\bar{u}^*$.

Proof: Define the estimation error as $\tilde{W}=W-\hat{W}$, then similar to [18, 23], one can solve the matrix equation (36) and verify from (37) that the fact $M=-P_2\tilde{W}+\upsilon$ holds, where $\upsilon=-\int_0^t e^{-\ell(t-r)}\varepsilon_{HJB}(r)\Xi^T(r)dr$ is a variable bounded by a positive constant ε_1 as $\|\upsilon\|\le\varepsilon_1$.

On the other hand, following a similar proof as given in [18, 23], if the vector Ξ in (35) is PE, we can verify that the matrix P_2 in (36) is positive definite, which means $\lambda_{\min}(P_2)>\sigma_2>0$ holds for any positive constant $\sigma_2>0$.

Select a Lyapunov function as $L=\frac{1}{2}\tilde{W}^T\Gamma^{-1}\tilde{W}$, then the derivative $\dot{L}$ can be calculated from (37)-(38) as

$$\dot{L}=\tilde{W}^T\Gamma^{-1}\dot{\tilde{W}}=-\tilde{W}^TP_2\tilde{W}+\tilde{W}^T\upsilon. \tag{39}$$

i) when $\varepsilon_{HJB}=0$, we know that $\upsilon=0$, such that (39) can be written as

$$\dot{L}=-\tilde{W}^TP_2\dot{\tilde{W}}\le-\sigma_2\left\|\tilde{W}\right\|^2\le-\mu L \tag{40}$$

where $\mu=2\sigma_2/\lambda_{\max}(\Gamma^{-1})$ denotes a positive constant. From (40), one can claim that $\tilde{W}$ will exponentially converge to zero. Therefore, we know $\hat{W}\to W$ for $\varepsilon_v=0$. In this case, the error between $\bar{u}^*$ in (23) and $\bar{u}$ in (33) is given as

$$\bar{u}^*-\bar{u}=-R^{-1}B^T\nabla\phi^TW+R^{-1}B^T\nabla\phi^T\hat{W}=-R^{-1}B^T\nabla\phi^T\tilde{W} \tag{41}$$

Hence, $\lim_{t\to+\infty}\left\|\bar{u}^*-\bar{u}\right\|=0$ holds, i.e. $\bar{u}$ converges to $\bar{u}^*$.

ii) when $\varepsilon_{HJB}\ne 0$, Eq.(39) can be represented as

$$\dot{L}=-\tilde{W}^TP_2\dot{\tilde{W}}+\tilde{W}^T\upsilon\le-\left\|\tilde{W}\right\|(\sigma_2\left\|\tilde{W}\right\|-\varepsilon_1) \tag{42}$$

Based on (42) and Lyapunov's Theorem, we can claim that

$\tilde{W}$ will converge to a compact set $\Omega_1 : \left\{ \tilde{W} \mid \left\| \tilde{W} \right\| \le \varepsilon_1 / \sigma_2 \right\}$, whose size is determined by the approximator error ε_1 and the excitation level σ_2. In this case, we recall (23), (33) to evaluate the error of the control policy, and have

$$\begin{aligned} \bar{u}^* - \bar{u} &= -R^{-1}B^T(\nabla\phi^T W + \nabla\varepsilon_v) + R^{-1}B^T\nabla\phi^T\hat{W} \\ &= -R^{-1}B^T\nabla\phi^T\tilde{W} - R^{-1}B^T\nabla\varepsilon_v \end{aligned} . \quad (43)$$

Hence, we can verify that

$$\lim_{t\to+\infty} \left\| \bar{u}^* - \bar{u} \right\| \le \lambda_{\max}(R^{-1}B^T)(\phi_M \left\| \tilde{W} \right\| + \phi_\varepsilon) \le \varepsilon_u \quad (44)$$

holds for a positive constant $\varepsilon_u > 0$. ◇

We can now present an online algorithm to derive the approximated solution of equation (22) for observer (17).

Algorithm 2-Online Adaptation for Optimal Observer

1: **Initialization**: Select initial condition $\hat{W}(0)$ and parameters Γ, ℓ for adaptive law (38);

2: **Start procedure**

3: **i) Online adaptation**: Collect $\hat{x}, y$ for $\phi(X)$, and online calculate P_2, Q_2, M and $\hat{W}(t)$ along (38) using any ordinary differential equation (ODE) solver (e.g. Runge-Kutta) for integration interval $t \in [t_i, t_{i+1}]$, $i \in N$.

4: **ii) Observer implementation:** obtain the correction term $\bar{u} = -R^{-1}B^T\nabla\phi^T(X)\hat{W}$, and implement the observer (17).

5: **iii) Continuation:** let $i := i+1$

6: **end procedure**

Remark 4: Theorem 2 shows that the suggested Algorithm 2 with online adaptation can guarantee that the approximated policy $\bar{u}$ in (33) converges to the optimal solution $\bar{u}^*$ in (23). Therefore, based on Theorem 1, the proposed observer (17) with online updating compensator (33) is stable.

Remark 5: As shown, the Algorithm 2 is implemented in an online manner, i.e. an offline iteration procedure is not needed. Moreover, the initial stable compensator is not assumed as we proved that the suggested adaptive law (38) can guarantee that $\hat{W}$ converges to its true value W under PE condition. Thus, the approximated policy $\bar{u}$ in (33) converges to the ideal optimal policy (23).

IV. SIMULATIONS

This section will present two simulation examples to validate the efficacy of the suggested methods.

A. Observer for Linear System

We consider a linear system

$$\begin{cases} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -0.16 & -0.56 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \\ y = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{cases} \quad (45)$$

This is indeed the linear system (1) with $A = \begin{bmatrix} 0 & 1 \\ -0.16 & -0.56 \end{bmatrix}$, $B = [0 \quad 1]^T$ and $C = [1 \quad 0]$. To design the optimal observer, we choose the weighting matrices as $Q = 100$, $R = 0.1$ to obtain satisfactory observer error convergence, then the standard observer solution of ARE (14) is obtained as $P^* = \begin{bmatrix} 25.14 & 3.15 \\ 3.15 & 0.7387 \end{bmatrix}$ by using the Matlab command 'care' or 'lqr'.

The initial conditions are $x(0) = [8 \quad 2]^T, \hat{x}(0) = [5 \quad 0]^T$. then simulation results of observer (2) with (14) and (15) are shown in Fig. 1, where the profiles of the plant state and the observer state are all illustrated. One may find that the observer tracks the derived system dynamics well when the optimal correction term is applied.

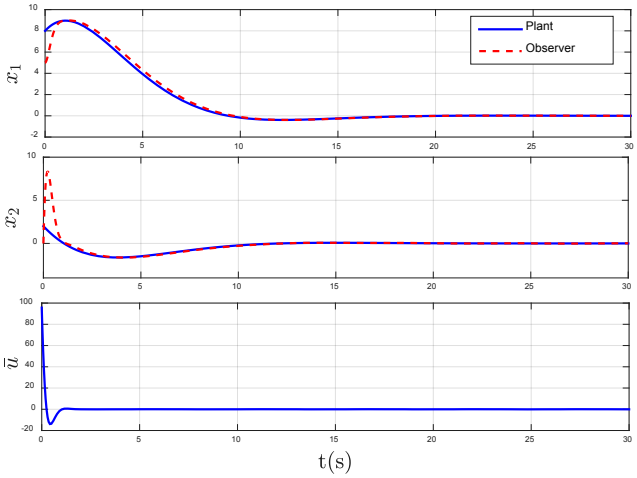


Fig.1 Observer performance of observer (2) with (14) -(15).

B. Observer for Nonlinear System

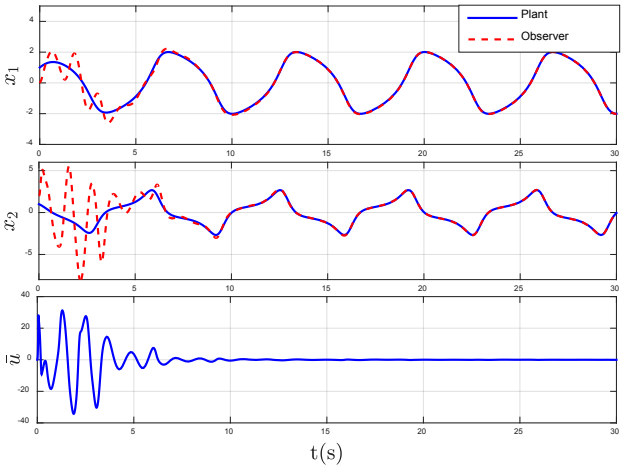
Consider the *Van der Pol* oscillator system given by:

$$\begin{cases} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} (-x_1^2 x_2) \\ y = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{cases} \quad (46)$$

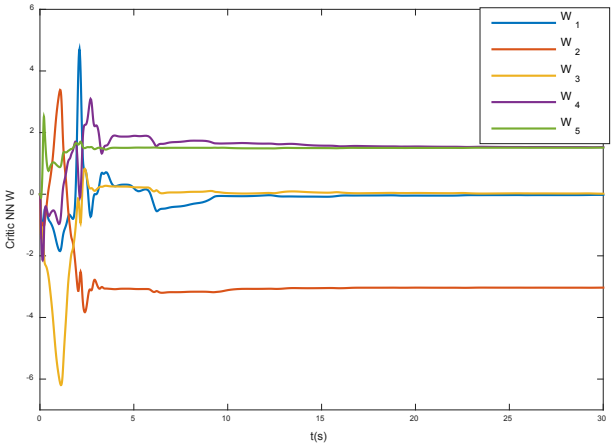
Then the observer (17) with the approximated policy (33) and adaptive law (38) is implemented with simulation parameters $\ell = 1$ and $\Gamma = 5diag([1,1,1,1,1])$. The regressor vector for approximating the cost function is $\phi(x) = [y, y\hat{y}, \hat{y}, \hat{y}^2, y^2]^T$. Moreover, the weighting matrices in the performance function (18) are set as $R = 0.1$ and $Q = 100$. The initial conditions are set as $x(0) = [1,1]^T, \hat{x}(0) = [0,2]^T$ and $W(0) = [0,0,0,0,0]^T$.

Simulation results are given in Fig.2, where the system state, the observer state and the compensation term are illustrated in Fig.2 (a), and the profiles of the critic parameters are shown in Fig.2 (b). One may conclude from Fig.2 (a) that the observer can converge to the given system state after a transient, and the correction term converges to zero. Fig.2 (b) also indicates that the suggested adaptive law (38) can achieve the

parameter convergence of the critic parameters.



(a) Profiles of observer state and correction term.



(b) Critic approximator parameter $\hat{W}$.

Fig.2 Nonlinear observer with optimal policy (33) and adaptation (38).

V. CONCLUSION

This paper addresses the optimal observer design by considering the duality of the observer design and the optimal tracking control. For linear systems, the optimal observer is reformulated as a linear quadratic tracking (LQT) control problem. Then the feedforward and feedback actions in the observer can be obtained by solving an ARE. For nonlinear systems, we represent the observer design as the output tracking control, where a compensation term is used to address the nonlinearities and to minimize a cost function. Then an offline policy iteration method is introduced to solve the optimization problem. Finally, we extend the idea of ADP to online solve the optimization equations, where a critic approximator is used to estimate the optimal value function, and a novel adaptive law is used to online update the unknown critic parameters. The convergence to the optimal solution is rigorously guaranteed. Simulation results are provided to illustrate the efficacy of the proposed observers.

REFERENCES

[1] D. G. Luenberger, "Observing the state of a linear system," IEEE Transactions on Military Electronics, , vol. 8, pp. 74-80, 1964.

[2] G. Ciccarella, M. Dalla Mora, and A. Germani, "A Luenberger-like observer for nonlinear systems," International Journal of Control, vol. 57, pp. 537-556, 1993.

[3] A. N. Atassi and H. K. Khalil, "Separation results for the stabilization of nonlinear systems using different high-gain observer designs," Systems & Control Letters, vol. 39, pp. 183-191, 2000.

[4] M. S. Mahmoud, Robust control and filtering for time-delay systems. New York,: CRC Press, 2000.

[5] B. L. Walcott and S. H. Zak, "Combined observer-controller synthesis for uncertain dynamical systems with applications," Systems, Man and Cybernetics, IEEE Transactions on, vol. 18, pp. 88-104, 1988.

[6] S. K. Spurgeon, "Sliding mode observers: a survey," International Journal of Systems Science, vol. 39, pp. 751-764, 2008.

[7] J. Na, G. Herrmann, X. Ren, and P. Barber, "Adaptive discrete neural observer design for nonlinear systems with unknown time-delay," International Journal of Robust and Nonlinear Control, vol. 21, pp. 625-647, 2011.

[8] V. Durbha and S. N. Balakrishnan, "New nonlinear observer design with application to electrostatic micro-actuators," in ASME 2005 International Mechanical Engineering Congress and Exposition, 2005, pp. 101-107.

[9] H. Kwakernaak and R. Sivan, Linear optimal control systems: Wiley-interscience New York, 1972.

[10] F. Lin, Robust control design: an optimal control approach vol. 18: John Wiley & Sons, 2007.

[11] Y.-C. Ho and A. E. Bryson, "Applied optimal control," New York: Hemisphere, 1975.

[12] C. P. Mracek, J. R. Clontier, and C. D'Souza, "A new technique for nonlinear estimation," in Proceedings of the IEEE International Conference on Control Applications, 1996, pp. 338-343.

[13] V. Durbha, S. N. Balakrishnan, and W. Dyer, "Target interception with cost-based observer," in AIAA Guidance, Navigation, and Control Conference and Exhibit 2006, 2006.

[14] P. J. Werbos, "Approximate dynamic programming for realtime control and neural modeling," in Handbook of intelligent control: Neural, fuzzy, and adaptive approaches, D. A. White and D. A. Sofge, Eds., ed: New York: Van Nostrand Reinhold, 1992, pp. 67-95.

[15] Y. Lv, J. Na, Q. Yang, X. Wu, and Y. Guo, "Online adaptive optimal control for continuous-time nonlinear systems with completely unknown dynamics," International Journal of Control, vol.89, pp. 99-112, 2016.

[16] J. Na and G. Herrmann, "Online adaptive approximate optimal tracking control with simplified dual approximation structure for continuous-time unknown nonlinear systems," IEEE/CAA Journal of Acta Automatica Sinica, vol. 1, pp. 412-422, 2014.

[17] F. L. Lewis, D. Vrabie, and V. L. Syrmos, Optimal control: Wiley. com, 2012.

[18] J. Na, M. N. Mahyuddin, G. Herrmann, X. Ren, and P. Barber, "Robust adaptive finite time parameter estimation and control for robotic systems," International Journal of Robust and Nonlinear Control, pp. 1-27, 2015 (In press).

[19] H. Modares and F. L. Lewis, "Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning," Automatic Control, IEEE Transactions on, vol. 59, pp. 3051-3056, 2014.

[20] D. Kleinman, "On an iterative technique for Riccati equation computations," Automatic Control, IEEE Transactions on, vol. 13, pp. 114-115, 1968.

[21] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," Automatica, vol. 41, pp. 779-791, 2005.

[22] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," Automatica, vol. 45, pp. 477-484, 2009.

[23] J. Na, X. Ren, and Y. Xia, "Adaptive parameter identification of linear SISO systems with unknown time-delay," Systems & Control Letters, vol. 66, pp. 43-50, 2014.

[24]G Besançon, I Munteanu, "Control strategy for state and input observer design", Systems & Control Letters, vol. 85, pp. 118-122, 2015.